

**But I'd rather have raisins! Exploring a hybridized approach to multimodal interaction in the case of a minimally verbal child with autism**

DOAK, Luran <<http://orcid.org/0000-0002-7934-5276>>

Available from Sheffield Hallam University Research Archive (SHURA) at:

<http://shura.shu.ac.uk/18473/>

---

This document is the author deposited version. You are advised to consult the publisher's version if you wish to cite from it.

**Published version**

DOAK, Luran (2018). But I'd rather have raisins! Exploring a hybridized approach to multimodal interaction in the case of a minimally verbal child with autism. Qualitative Research. (In Press)

---

**Copyright and re-use policy**

See <http://shura.shu.ac.uk/information.html>

# **But I'd Rather have Raisins! Exploring a Hybridized Approach to Multimodal Interaction in the Case of a Minimally Verbal Child with Autism.**

## **Abstract**

*This article explores a 'hybridized approach' to multimodal research drawing on video data of classroom communication involving children diagnosed with Autism Spectrum Disorder. The focus is a short video of 'Luke', aged six, who at snack time declines to request an available food item (carrot, tomato, or apple) with the available Picture Exchange Communication System (PECS); instead deploying embodied, idiosyncratic communication including gaze, vocalisation and object manipulation to request raisins. The article explores the potential of a hybridized approach for understanding Luke's communicative competencies which draws upon the theoretical perspectives of Ethnography of Communication, Conversation Analysis and Multimodal (Inter)Action Analysis; and uses two forms of multimodal transcription (the multimodal matrix and annotated video stills). It is argued that each tradition brings distinct affordances to our understanding of this short interaction and that together they can permit inferences which would not have been possible working with one approach alone.*

**Keywords:** *Multimodality, Multimodal (Inter)Action Analysis, Conversation Analysis, Ethnography,*

**AAC, Autism**

## **Introduction**

The relatively new field of 'multimodality' encompasses a wide proliferation of approaches to research including social semiotics, systemic functional analysis, conversation analysis, geo-semiotics, Multimodal (Inter)Action Analysis, multimodal ethnography, multimodal corpus analysis and multimodal reception analysis; each with their own epistemological and methodological

commitments in the study of communication. Additionally, many 'multimodal' studies are primarily embedded in the languages of their own established disciplines such as education, advertising, architecture and film studies; which can present a challenge in terms of establishing common ground and shared understandings of multimodality in the context of domain-specific vocabularies (O'Halloran et al., 2012). Attempts have nevertheless been made to establish common ground in multimodal research. According to Jewitt et al. (2016) these include the recognition that human interaction is undertaken with a wide range of semiotic resources which realise different communicative work in a multimodal ensemble because of the affordances and constraints of their materiality; that language should not be *a priori* privileged over other modes nor should 'non-verbal modes' should not be presumed to play an orbital or supporting role to language; and that it is important to analyse how communicators select and orchestrate semiotic resources to produce a 'multimodal whole'. This commonality raises the question of whether it is possible to draw upon concepts from diverse multimodal perspectives to form a 'hybridized approach' to multimodal analysis.

Jewitt (2009) argues that whilst different approaches to multimodality have evolved to attend to particular aspects of multimodal meaning-making, boundaries between perspectives 'will be contested and remade .... [and] provide useful opportunities to cross and transgress, to rethink and to collaborate across' (p.29). At the same time, there is a need for reflection on the degree of compatibility between the multimodal concepts and the theoretical and methodological frame into which they are integrated (Jewitt et al. (2016). This paper considers the value of a 'hybridized approach' to multimodal analysis, combining elements of Ethnography (specifically, Ethnography of Communication), Conversation Analysis and Multimodal (Inter)action Analysis. This methodological exploration will be applied to the communicative competencies of a minimally verbal child with Autism Spectrum Disorder (ASD), an area of inquiry that requires careful attention to communication beyond language. In the following, I will briefly introduce each of these elements separately before considering how they could be combined.

## *Ethnography*

An ethnographic approach to classroom research tends to involve direct and sustained contact with participants in their everyday lives using a wide range of methods including participant observation, fieldnotes, audio and videorecordings, interviews and the collection of photographs, artefacts and contextualising documents; with the aim of producing a rich qualitative account which values both *emic* and *etic* perspectives. The proposed framework draws specifically upon concepts derived from *Ethnography of Communication (EoC)* (Hymes, 1972); which explores the nexus between language and culture. It seeks firstly to identify the *speech community* (a group whose members have significant commonality in how they use, value or interpret language); and then to elucidate the nature of these shared practices. Specifically, it addresses the issue of *communicative competence* within the community: what does a speaker need to know to communicate appropriately within the speech community, and how do they learn to do so? This question goes far beyond interactional competence in the linguistic sense, asking what may be said, when, how and by whom. The concept of 'speech community', however, is not straightforward: a group may comprise multiple overlapping and interacting communities, and an individual may simultaneously identify (to varying extents) with more than one community. Even within one identified 'speech community' there is variation in the resources available to individual members, with Saville-Troike (2008) noting that 'different subgroups of the community may understand and use different subsets of its available codes' (p.41).

EoC uses three units of analysis: the *communicative act* (an observable behaviour which seems to contain a speech function); the *communicative event* (a series of interconnected communicative acts which are bound together by a topic or purpose); and the *communicative situation* (the context within which the event unfolds with its associated interactional norms, expectations, rituals and prohibitions). This wider contextualisation is considered entirely compatible with more detailed microanalyses of communicative acts and events, which 'are in a necessary complementary relationship to one another if an understanding of communication is to be reached' (Saville-Troike, 2008:106). The EoC framework thus provides the possibility of contextualising small, fleeting fragments of interaction by locating them within wider understandings of the classroom

communicative culture and the beliefs and values attached to (for example) the relative privileging of different modes.

More specifically in relation to 'disordered' communication, Kovarsky et al (1988) proposed what they termed an 'Ethnography of Communication Disorders' (ECD) which drew upon the field methods and analytic tools of EoC to explore the relationship between language, culture and clinically identified difficulties in communication. Reflecting on the contribution of ECD some years later, Kovarsky (2016) argues that ECD has enhanced clinical understandings of communication disorders in at least three ways. Firstly, it has challenged the traditional epistemology of communication disorders (framed by a positivist paradigm which values objective and quantifiable measures of 'progress') to recognise also the clinical significance of understanding the feelings, rationale and emic perspective of the 'client'. Secondly, ethnographic observation of the interactional patterning of therapy sessions with clients illuminated and problematised features previously considered unremarkable such as the 'necessary [adoption of] roles as competent expert and incompetent patient in order for therapy to proceed in an orderly and efficient fashion' (Simmons-Mackie and Damico, 1999: 313). Thirdly, it has argued that (contrary to traditional understandings of communication disorders as demonstrable entities evidenced by standardised test scores) 'communication disorders are brought into existence by their social and cultural consequences through inter-subjective experiences of stigmatization, marginalization, and a diminished sense of place and identity' (Kovarsky, 2016:13).

In a similar vein, Solomon (2008) argues that ethnography can provide a useful counterpoint to the clinical view of disordered language as a 'disembodied cognitive process awaiting remediation' (p.150); by insisting on the study of children communicating *in situ* as members of families and communities where they are 'socialised into sociocultural competence' (p.150) and where patterns of language use are always linked to particular cultural practices. Ochs et al (2004), use an ethnographic approach to contest decontextualized concepts in diagnostic criteria such as perceived deficits in interpersonal perspective taking; arguing that any 'interpersonal' exchange unfolds in a sociocultural setting of organised practices, roles, institutions beliefs and knowledge

which must be properly understood. Such studies suggest that adopting an ethnographic perspective on the communication of children with autism serves as an important reminder that 'while social functioning needs to be understood as a general domain of ability, it also needs to be examined as an on-line, real-time process involving knowledge of historically rooted and culturally organized social practices' (Ochs et al, 2004: 157). Another approach that attends to real-time communication, and has direct relevance to communication disorders such as ASD, is Conversation Analysis.

### *Conversation Analysis*

Conversation Analysis is a methodological approach to the study of everyday talk in interaction. Interactions are audio or video-recorded, systematically transcribed and analysed in order to make visible the normally taken-for-granted '*machinery of conversation*' (Liddicoat, 2011). Transcription often uses the 'Jefferson system' which in addition to transcribed speech provides for symbolic notation of features such as pauses, eye gaze, prosodic features, laughter and overlap (Jefferson, 2004). A core premise is that contributions to interaction are simultaneously *context-shaped* and *context-renewing*: that is, any given utterance is constrained by the limited range of potentially relevant next actions suggested by the previous utterance, and in turn contributes to the sequentiality of the interaction by setting up its own limited range of potentially relevant next actions for the next interactant (Heritage, 1984).

Based on the premise of sequentiality, CA has elaborated on how interactants realise certain features of conversation including openings and closings, turn-taking, adjacency pairs, preference organisation and repair. For instance, turn-taking is structured around the *Turn Constructional Unit* (TCU); which denotes a recognizably complete and meaningful contribution in the ongoing talk (Sacks et al., 1974). Towards the end of a TCU comes a *Transition Relevance Place* (TRP) which the speaker may subtly indicate by changes in syntax, eye gaze, intonation and/or prosody; and it is in the TRP that a change in speaker becomes a legitimate next action (Sacks et al., 1974). Related to this, an 'adjacency pair' denotes a pair of TCUs which belong together; the first of which has a normative force in determining the content of the second (Heritage, 1984). Commonly-seen types

include greetings (requiring a return greeting); terminal adjacency pairs (requiring return of 'goodbye'); invitation/ offer adjacency pairs (requiring a response); assessments (evaluations of a situation under discussion requiring assent or dissent); complaints (requiring excuse or remedy); information (requiring acknowledgement) and questions (requiring an answer). Failing to provide the expected completion would be an *accountable* action requiring repair, since participants in interaction continually attend to the matters of mutual understanding.

CA also proposes the concept of *preference organisation*. Atkinson and Heritage (1984) note that certain *preferred* actions in conversation (such as agreeing with an assessment or accepting an invitation) are performed immediately and without delay; whilst other *dispreferred* actions (disagreeing or declining) tend to be accomplished with extra conversational work. This might include a *hedge* ('I dunno'), a *warrant* ('I'd love to but I'm so busy right now ...'), a *token* ('uhm', 'uh', 'well') or *weak agreement* ('Yeah I suppose that might be it'). The purpose of this extra work is to mitigate the possible effects of a dispreferred action which could otherwise be perceived as rude or hostile (Goodwin and Heritage, 1990).

The early literature on CA has been accused of giving undue primacy to the role of verbal speech in communication (Erickson, 2010); both in its data collection methods (primarily audio-recordings) as well as its transcription practices which tended to focus on speech, eye gaze and 'non-lexical soundmaking' (Thomas, 1987) such as 'sigh', 'in-breath' and laughter. Whilst analysis of embodiment in interaction was certainly not absent from the early literature (see for example Enninger, 1987; Goodwin & Goodwin, 1986; Sigman, 1987); Nevile (2015) identifies a significant '*embodied turn*' in CA literature taking place from 2001 onwards which characterised by increased exploitation of video-recording technologies to enable visual representation and analysis of the role of the body in social interaction.

Subsequently, a body of multimodal research in the CA tradition has developed which is sometimes referred to as 'multimodal interaction' research (not to be confused with the similarly named but theoretically distinct Multimodal (Inter)action Analysis (Norris, 2004) which is discussed separately later). For instance, Mondada (2016) argues that CA is well placed to bring 'careful and

precise attention to temporally and sequentially organized details of actions that account for how co-participants orient to each other's multimodal conduct, and assemble it in meaningful ways, moment by moment' (p.340). By way of example, the same author undertakes analysis of the unfolding of a surgical theatre procedure using conventional Jefferson transcription supplemented with photographs and additional notation symbols to facilitate the insertion of verbal descriptions of embodied action (Mondada, 2011); noting 'a complex web of situated collective multimodal actions' (p.224) where multiple parallel streams of action (some compatible, some mutually exclusive) are fluidly co-ordinated through multimodal alternating and sequencing procedures. Stivers and Sidnell (2005) draw a distinction between the vocal/aural and visuospatial modalities; arguing that the interactional work undertaken by one modality may support, extend or modify that which is undertaken by the other and that both provide important resources in the collaborative production of emergent turns-at-talk'. (p.15). Goodwin (2011) uses traditional CA transcription with arrows to linked line drawings of participants to explore how a man with aphasia and only three spoken words can nevertheless participate successfully in complex interaction through a process which the author names *cooperative semiosis*; observing how the aphasic participant can 'vastly expand his repertoire as a speaker by sequentially typing to the particulars of the complex talk and language structure of his interlocutors' (p.186). Elsewhere, Goodwin (2007) uses the same transcription approach to explore what he terms *embodied participation frameworks* (the way in which participants physically orient their bodies toward each other and the subsequent implications of this framework for the affective, cognitive, gestural and artefactual alignment of the interaction that takes place within it). Lerner et al (2011) demonstrate with the use of video stills how a sixteen month old infant is able to make use of the 'activity context' (the sequential structure of the caregiver's actions as she feeds another child) as a framework for the composition and placement of her own (pre-lingual, embodied) demands for food. This selection of studies, although not comprehensive, is intended to give a flavour of how CA has engaged with the role of the body in the sequential organisation of the 'machinery of conversation' (Liddicoat, 2011).



CA therefore has affordances in the study of communicative competencies through the systematic study of the sequential organisation of interaction. This has the potential to challenge and disrupt conventional understandings of individual 'deficit' in children with atypical communication (Muskett et al., 2010) by exploring the functionality of an action (however idiosyncratic) within the unfolding sequence, and uncovering competencies which might otherwise have been overlooked. Finally, *Multimodal (Inter)action Analysis*, an approach for exploring the intensity and complexity of multiple modes could be a useful for the study of communicative competencies in minimally verbal children with ASD.

### *Multimodal (Inter)action Analysis*

Multimodal (Inter)action Analysis (Norris, 2004) is a framework for the analysis of multimodal interaction which is theoretically located in the interface between interactional sociolinguistics (Gumperz, 1982); mediated discourse analysis (Scollon, 2001); and multimodality (Kress & Van Leeuwen, 2001). This tripartite heritage gives the framework a distinctive approach to the study of multimodal interaction which focuses on real-time interaction through multiple modes which is always deeply embedded in the geosemiotic world of artefacts and mediational tools. The strong emphasis on the inseparability of multimodal human (inter)action from the surrounding material world is reflected in Norris' preference for annotated video stills as the primary means of transcription. It could also be said to take a more wide-angled lens to the study of interaction than the Conversation Analytic focus on immediate interactions at sequential level; instead choosing to embrace analysis of how features of the surrounding environment (such as background noise, music, furniture and passers-by) may influence the unfolding exchange.

Norris' MIA framework takes as its analytic focus the continual intersection of diverse modes in an interaction and how these may serve to foreground or background the concerns of the actors. Interactants undertake '*higher-level actions*' which are clearly bracketed by an opening and closing. These higher-level actions in turn are composed of chains of '*lower-level actions*' (successive shifts in eye gaze, posture, proxemics, language, head movements, and engagement with artefacts). Higher-

level actions may be brought to the foreground of our continuum of attention by either high *modal complexity* (where many modes are oriented towards the realisation of the same higher-level action) or high *modal intensity* (where one mode is particularly salient in that the performance of the higher-level action depends upon it, such as the pivotal role of voice during a telephone call). The concept of 'attention' as used by Norris explicitly rejects the idea of actions as a transparent window into cognitive processes: as she cautions, 'the actual experience and the expression of the experience should not be viewed as a one-to-one representation and may be as diverse as to contradict each other' (p.4). Nevertheless, she maintains, it is possible through detailed qualitative analysis of the modal intensity and/or complexity of observable behaviours to make suggestions about the relative positioning of multiple concurrent higher-level actions on a participant's continuum of awareness/attention.

The MIA framework could be helpful in viewing minimally verbal participants as competent, agentic communicators who actively deploy multiple modes in ever-changing configurations of varying intensity and complexity just as verbal communicators do. This is facilitated by Norris' preferred transcription method (annotated video stills) which consciously de-privileges language in order to foreground the role of non-verbal modes such as proxemics and posture. I will now consider how these elements could be combined to form a hybridized approach to multimodal analysis.

### *A Hybridized Approach to Multimodal Analysis*

In this study, elements of the three approaches described above are drawn together in the analysis of a short piece of classroom video-recorded data. Kress (2011) speaks of the possibility of 'complementarity' between ethnography and forms of multimodal analysis, based on the question of 'reach' (p.241): what does a theory or methodology do well or not do well for a given research question, and where does its 'reach' run out? From the ethnographic perspective, data collected from a wide range of sources beyond the immediate transcription can usefully contextualise the subsequent microanalysis. This 'rich backstory' (Flewitt, 2011:307) provided by ethnography is

considered fundamental to this analysis: the video-recorded event (snack time) does not occur in a contextual vacuum but rather within an established 'communicative situation' (snack time) which in turn draws on pedagogical beliefs and practices in special education to inform its enactment.

However, the admissibility of ethnographic contextualising detail alongside multimodal microanalysis has been also contested: McHoul et al. (2008) note a 'sequential purism' in CA which considers only context which is empirically evidenced and invoked in participants' talk to be analytically relevant. Maynard (2006:83) argues for a 'limited affinity' between CA and ethnography; with admission of the wider-than-sequential context only where it is procedurally consequential in the unfolding interaction. Nonetheless, a multimodal microanalysis without contextualising ethnographic detail could obscure imbalances of interactional power between participants (particularly relevant in the case of participants with learning disabilities): Svennevig et al. (2005) argue this can 'direct analytic attention away from partially shared resources, misunderstanding and unequal rights to define the procedures to be employed' (p.11). Ethnography of Communication is particularly well-placed to reflect on questions such as who decides what may be said; how it may be said; who has access to which semiotic resources; and which modes are privileged above others. For instance, Moerman (1988), in his call for a 'culturally contexted conversation analysis' (p.6) states:

[CA] has much to learn from [Ethnography of Communication's] consistent recognition that societies differ in their ways of speaking both from one another and internally, and from the prominence that it gives to the historical background, investigated contexts, and rich cultural meanings of speech events. (p.11)

The hybridized approach in this paper draws from CA the proposition that a closely detailed transcription, which captures the temporal unfolding of sequential interaction, is invaluable in foregrounding the functionality of atypical communicative acts, and has consequently influenced the exploration with transcription. In what follows, the paper draws upon and appropriates the concepts of CA including sequentiality and features of conversational organisation, such as turn-taking and preference organisation, where these facilitate analysis of the present data.

The approach further draws upon concepts derived from Multimodal (Inter)Action Analysis, specifically modal intensity and complexity. Whilst a multimodal approach to CA has evolved to attend specifically to the sequential functionality of multimodal actions in interaction; MIA brings a different, and perhaps complementary, focus on how dynamic fluctuations of modal complexity and intensity are used to foreground the participants' interactional concerns. Further, Norris' insistence on the de-privileging of language (both theoretically and methodologically with visual transcripts) is a useful counterpoint to the historically logocentric tradition of CA and contributed to the decision to use annotated video stills as a means of transcription. I will consider next the relevance of this for researching ASD.

### *ASD and communication*

ASD is medically understood as an impairment of social interaction featuring repetitive and restrictive patterns of interests and behaviours; sensory processing difficulties; and deficits in language and other communication skills (APA, 2013; WHO, 1992). Approximately 30% of people with a diagnosis of ASD are non-verbal or minimally verbal (Tager-Flusberg et al., 2013); minimally verbal denoting no more than 20-30 spoken words (Kasari et al., 2013). Augmentative and Alternative Communication (AAC) is recommended to ensure that minimally verbal children do not develop a pattern of communication failure (Prizant et al., 2003); with approaches such as Picture Exchange Communication System (PECS), Makaton signing, or speech-generating devices (SGDs) being commonplace in UK special education (Sheehy et al., 2009; Roulstone et al., 2012). This section briefly reviews the multimodal literature on minimally verbal AAC users from the ethnographic and CA perspectives; although it is acknowledged that this has also been usefully explored from the perspective of social semiotic multimodality (Dreyfus, 2006; Flewitt et al., 2009).

A number of studies have used ethnographic methods to study the classroom communication of minimally verbal children. Using an ethnographic case study approach, Mellman et al. (2010) observed students being communicatively disabled by AAC inaccessibility (their device was left on a counter out of reach); limited staff training; staff attitudes; missed opportunities to

programme useful vocabulary relating to school life; and the devaluing of social interaction with peers. They additionally observe that many interactions relied on gesture, facial expressions and non-verbal vocalisations which were not always given the same recognition as AAC-mediated communication. In the study by Flewitt et al. (2009), ethnographic video case studies of preschool children were undertaken across multiple settings (home and two educational environments). They observed significant differences in communication practices and expectations in each environment, with embodied, idiosyncratic communicative competencies being more valued in the home setting and the more 'inclusive' educational setting with the specialist setting prioritizing formal augmented communication such as Makaton and PECS. The foregrounding of the environmental contribution to communicative practices are therefore a significant potential affordance of ethnographic studies, as teachers may be unaware of the extent to which school timetabling, routine and expectations disable communication which is happening in more relaxed environments

CA has also contributed to the literature on minimally verbal communicators; with a number of studies examining the embodied communication of minimally verbal students in the absence of AAC. For instance, Korkiakangas et al. (2013) use video data from a classroom interaction to examine the interactional role of the manipulation of material objects; Dickerson et al. (2007) analyse the interactional significance of physically tapping on presented items; and Stribling et al. (2007) use CA to reframe 'echolalia' (repetition of previous utterances) as a productive form of interactional work. Muskett et al. (2013) argue that in the case of participants with communication disorders it may be essential for CA to adopt a more multimodal orientation than usual in order to facilitate analysis of the orderliness of the participant's use of 'multiple semiotic resources including, but not limited to, talk' (p.837).

CA has also been used to examine AAC usage by participants with a variety of communication disorders. For instance, Bloch et al. (2004) demonstrate how two AAC users attempt self-repair of communication problems via their devices, concluding that qualitative AAC studies can reveal how embodied and technologically aided modes co-exist in a largely complementary manner. Similarly, Clarke et al. (2013) examine how an AAC user switches his eye gaze from his device to his

interactional partner as part of the speaker transfer negotiation; whilst Wilkinson (2013) observes an AAC user supplementing his speech with iconic gestures which contribute semantic meaning to the interaction but also accomplish social actions such as answering or repairing. Engelke et al. (2013) argue that CA is valuable to AAC insofar as it locates communicative success (or failure) in the collaborative and co-constructive activities of both the user and their communicative partner; and that such detailed microanalysis of this ongoing interactional negotiation can have important clinical implications by improving therapy programs and device design. Thus a body of work already exists on communication disorders from ethnographic and CA perspectives. This paper will build on this work with the hybridized approach that blends elements from each together.

### *Value of the Hybridized Approach for Exploring Minimally Verbal Communication*

Taken together, the three approaches outlined can offer distinct yet complementary contributions to our understanding of the idiosyncratic, atypical communication practices of a minimally verbal child. From ethnography, it is possible to contextualise fleeting instantiations of classroom communication within classroom, school and wider pedagogical concerns. The tools of CA can facilitate the identification and analysis of how minimally verbal interactants sequentially organise their interaction through multiple modes to enable turn-taking, repair of mishearings or misunderstandings, and the execution of preferred and dispreferred actions. Finally, Multimodal (Inter)Action Analysis considers how minimally verbal communicators actively orchestrate fluctuations in modal intensity and complexity to purposefully foreground and background their interactional concerns. In its appropriation of conceptual tools from three approaches, the present study is guided by the pragmatic question posed by Rampton et al. (2002): 'How do we need to adapt or hybridize these methods in order to say useful things about the practical problems on hand?' (p.375). I will start with considering transcription.

### *Approaches to Transcription*

A minimally verbal participant could be misrepresented as unresponsive or communicatively incompetent by transcription practices which fail to capture idiosyncratic, multimodal communication. This warrants critical reflection on the affordances and constraints of different transcription methods, with two of the three perspectives drawn upon (CA and MIA) having established transcription conventions. CA traditionally uses the Jeffersonian notation system (Jefferson, 2004); which provides a highly standardised approach to symbolic transcription of human interaction and places a high degree of emphasis on accurate transcription of the temporal, sequential unfolding of the interaction. Since CA originally developed from a corpus of primarily audio-recorded data, it's focus has been on transcribing the spoken word (but also other vocalisations, including in/out breaths and laughter); although more recently, CA has placed greater emphasis on transcribing multimodal communication through (for example) Jefferson transcriptions juxtaposed with video stills (Korkiakangas et al., 2014; Korkiakangas, 2018); the development of a set of extended conventions for transcribing embodied communication (Mondada, 2014); and Jefferson transcription combined with arrows linking to line drawings of relevant moments (Goodwin, 2011).

In contrast, MIA transcription intentionally problematises the presumed centrality of speech by choosing annotated video stills as the primary 'transvisual' and basis for analysis. As Norris (2004) argues, 'the prominence of spoken language is generally taken for granted in the field of discourse analysis, making it essential in a multimodal analysis to de-emphasize spoken language' (p.65). Norris does this as follows: speech is transcribed initially using Jeffersonian transcription, whilst sequences of shifts in other modes (gaze, gesture, posture, proxemics) are identified using series of extracted and time-stamped video stills for each mode. Finally, a transvisual is assembled to represent the overall interaction as clearly as possible, with a selection of chronologically-arranged video stills representing important interactional moments overlaid with a range of annotations. These may include arrows to indicate direction of movement and fragments of speech which are represented with a strong visual dimension to the text (e.g., curved text denoting variations in intonation; size and boldness indicating pitch; and physical space between pieces of text denoting the extent of gap or overlap).

In this paper, having reflected on the affordances of these established transcription conventions, the decision was taken to adopt neither in their entirety; instead preferring to match the hybridized approach to analysis with a hybrid two-stage approach to transcription consisting of a multimodal matrix (Fig.2) followed by annotated video stills (Fig.1) which would effectively illustrate the (atypical, minimally verbal) communicative competence of Luke. Multimodal matrices, which are more typically favoured in other multimodal perspectives such as social semiotics (Flewitt, 2006; Lancaster, 2007; Taylor, 2012) were useful at the analytic stage as they provided a frame for the temporary disaggregation of complex multimodal orchestrations and elucidating the contribution of individual modes to the overall *Gestalt*. In analytic terms, it draws attention to the contribution of less obvious modes, such as proxemics and posture, that might not be foregrounded on first viewing: the structure of the matrix frame ensured that they received equal analytic attention to other, more immediately salient modes, and mitigated against the risk of automatically privileging speech. The matrix also permitted detailed analysis of the sequentiality and temporal organisation of the exchange which is comparable to Jefferson transcription as it is chronologically ordered with time indicated in the far left column (see Figure 2); although Mondada's (2014) multimodal extension of the Jefferson system (as frequently used in multimodal approach to CA) achieves an even closer level of microanalysis with symbol notation of an action's preparation, apex, and retraction. As compared to Mondada's (2014) proposal of adding yet more symbolic notation conventions to an already heavily symbolised system, in the present paper, the (slight) compromise on microanalytic detail was considered justifiable: the matrix offered the combined affordances of a good level of sequential, time-annotated transcription, with a high degree of immediate readability for the uninitiated in CA.

The construction of the multimodal matrix was then followed by the (re)telling of the story of the exchange using time-stamped video stills, which draws loosely upon Norris' (2004) approach to transcription but keeps overlaid annotations minimal and includes instead a brief vignette-style commentary under each image. Video stills have particular affordances: they capture aspects of surrounding classroom layout and furnishing which may become relevant to the interaction, better



illustrate embodied interaction compared with verbal descriptions of a participant's physical movements, and situate the student in an interaction with a partner who is (ideally) also depicted in the video still in order to illustrate their physical and affective orientations towards each other. To 'tell the story' of Luke's multimodal competence, selected video stills or line drawings of moments from the (verbal) transcript did not seem sufficient to represent the spatial unfolding of a multimodal interaction where embodied actions are pivotal; thus annotated video stills have been used throughout. An advantage of the video stills is a high degree of 'readability' of the transcript, as audiences with no prior experience of multimodal transcription can easily follow the unfolding of the exchange. The issue of readability can be paramount in building dialogue with classroom practitioners and Speech and Language Therapists, when considering the differences between speech functions and vocabulary repertoires represented in AAC provision, and those which are demonstrably important to AAC users in their multimodal communication.

In sum, the decision to use two-fold transcription, although time-consuming, seeks to capture Luke's subtle, idiosyncratic, and unconventional communicative competences, and to enable detailed analysis of both sequentiality, and modal intensity and complexity, whilst situating the interaction in a broader ethnographic context.

## ***Methodology***

### *Context*

This paper draws on research undertaken in a classroom in a Special School in the Midlands of England. The class had a total of five students who ranged from five to seven years old, all with diagnoses of ASD and all minimally verbal (ranging from a few words to no spoken language). The classroom was staffed by one teacher and two teaching assistants. The study aimed to explore how the children made meaning as they went about their everyday lives, whether using AAC strategies or idiosyncratic embodied communication. Both PECS and Makaton signing were used and encouraged in this classroom; with student target-setting frequently referencing progress in one or both methods. My role as researcher in the classroom was part observer, part participant: some of my

time was spent on video-recording interactions with a small hand-held camera or taking notes; at other times I actively engaged with students or assisted Teaching Assistants with jobs such as tidying and supervising in the playground.

### *Participants and Ethics*

Jane is an experienced Teaching Assistant who has worked at the school for many years. She is a fluent Makaton signer and is also very familiar with PECS. Luke is six years old and was diagnosed with ASD and Global Developmental Delay aged three. He is developing some limited single word speech, knows a number of basic Makaton signs, and can use symbol cards to express his wants and needs when the symbols he requires are available. He very much enjoys social interaction using idiosyncratic embodied strategies such as gaze, touch, gesture and vocalisation.

Ethical considerations are particularly important when research involves children with learning and communication difficulties which may prevent them from verbally voicing concerns about the research. The study followed Nind's (2008) suggestion of proxy consent combined with an ongoing process of inferring the child's 'assent' to the research by reading their embodied responses to the presence of the researcher and the video camera; alongside consultation with classroom staff about the interpretation of such responses. Written consent for the research was obtained from the school, the classroom staff and the children's families; and the project was carried out in line with the BERA Ethical Guidelines for Educational Research (2011).

### *Data*

The study made use of ethnographic data collection methods although does not lay claim to being a full, immersive ethnographic study (Green and Bloome, 2004). Data was collected using observation and fieldnotes; video-recording of classroom interactions; photographs of classroom artefacts implicated in communication; collection of documents referencing classroom communication practices and pedagogy; audio-recorded interviews with staff and parents and a daily reflexive diary on the part of the researcher.

### *Transcription*

As noted above, transcription was undertaken using both a multimodal matrix and annotated video stills. The matrix involved repeatedly watching the short video clip in order to systematically examine each participants' use of speech, vocalisation, AAC, eye gaze, facial expression, gesture, object manipulation, proxemics (use of space), posture and haptics (use of touch). The sound was muted during analysis of modes such as posture and proxemics in order to focus analytic attention; and the video was at times watched in slow-motion or advanced frame-by-frame in order to establish the precise chronological ordering of events. The matrix is designed to be read chronologically by scanning from left to right to ascertain what each participant was doing at that point in time; or alternatively to use the colour coding of the modal groupings to identify how (for example) the postural and proxemic shifts of one participant influenced those of the other. The total matrix transcription of the video clip (which lasted 42 seconds) was five pages long, and the fourth page (which transcribes a sequence of particular analytic interest) is shown in Figure 2. Notational conventions were kept to a minimum, with ! and ? at the end of an utterance where a question or exclamation was apparent from intonation, syntax and/or context including accompanying non-verbal modes.; and with ... denoting a pause of any length (it was not considered necessarily to distinguish between pauses and micropauses as in the Jefferson system because the length of the pause is evident from the positioning of the utterance or act on the matrix).

The data was then transcribed again using annotated video stills. This transcription followed Norris in some respects (time-stamped video stills of selected interactional moments were arranged in chronological order and annotated in order to illustrate the unfolding interaction); but also differed in some respects (for instance, in the interests of readability text was printed in consistent size and font, which left the video still relatively unobscured but incurred the loss of transcribed intonation, pitch and prosody). Similarly, not every change in posture, proxemics, gesture or eye gaze was annotated in order to avoid obscuring the image. Spoken words or utterances were contained in speech bubbles whilst Makaton signs were placed in inverted commas near the hands

of the signing interactant. Notational conventions were minimal and consistent with their use in the matrix, and a short narrative description of each picture was placed underneath. The video still transcription in its entirety is represented in Figure 1.

### ***Case Study: But I'd Rather Have Raisins!***

In this case study, I will describe Luke's participation in snack time, an event which took place twice daily in this classroom, in a very standardised format. During the snack time, a C-shaped table was used, with the staff member leading snack time sitting on one side and the five students sitting around the other side of the table. This seating arrangement facilitated the enactment of snack time as the staff member could turn and physically realign themselves to face each student in turn with the snack tray (a large tray with four compartments to contain different snack items on offer).

When the snack tray was placed before a child, it would be accompanied by a PECS folder with laminated symbols representing the available items affixed to the front cover. It was a very consistent expectation that the child would lift the symbol for their desired item and hand it to the teacher to indicate their request. The teacher would then encourage them to verbalise the request and/or perform the Makaton sign for the item. When the item was given, the child would be prompted to perform the Makaton sign for 'thank-you' as a PECS symbol was not provided for this purpose. The tray and PECS folder would then pass to the next student, often rotating two or three times around the table until all the snacks had been distributed. From the perspective of Ethnography of Communication, snack time can be conceptualised as a 'communicative situation'. As Saville-Troike (2008) notes:

[it] maintains a consistent general configuration of activities, the same overall ecology within which communication takes place, although there may be great diversity in the kinds of interaction which occur there. (p.23)

My repeated observations of snack time revealed it being performed in a routinized format twice daily, and that there were certain shared expectations of how communication should be performed:

it took place in consistently designated times of day, and had physical artefacts associated with its enactment. Children were familiar with the PECS symbols as well as the expectations of how and when to use them, and it was relatively rare that any physical prompting was required. It was also clear that children were aware that the expectant pause when the teacher held up the symbol card indicated that they should attempt to express the choice in another mode (through spoken language or Makaton signing); and although children varied in their ability to produce spoken or signed language they would typically attempt one or the other. Thus the staff and children in this class formed a 'speech community' with a shared understanding of when PECS, Makaton, speech and embodied communication could and should be deployed in the various activities of the day. Some structured activities (such as lunchtime, snack time, and morning and afternoon group time) prioritised formal symbolic communication such as PECS, Makaton and speech whilst other activities, such as Intensive Interaction, privileged embodied communication such as facial expression, gaze, and vocalisation in playful, non-verbal exchanges designed to encourage reciprocity and mutual engagement. Nevertheless, this was not one homogenous 'community' with equally distributed resources. As Saville-Troike (2008) argues:

Within each community or complex of overlapping and interacting communities there exist a number of different language codes and ways of speaking available to its members ... it is very unlikely that any individual is able to produce the full range; different subgroups of the community may understand and use different subsets of its available codes. (p.41)

Whilst in the classroom, there were shared communicative practices to justify conceptualising it as a 'community', it was also the case that staff could orient to an alternative 'community' of fluent English speakers by a form of 'code-switching' when they spoke rapidly to each other without AAC support. It is difficult to ascertain whether children possessed a form of peripheral membership or participation in this community: the extent of each child's receptive understanding of fluent English was unclear and their expressive repertoire ranged from a few single words to none. (Although, as Dreyfus [2006] argues, minimally verbal communicators are thoroughly embedded in a 'transmodalised' speaking environment where their modes are often 'translated' into words.)

Similarly, membership of the 'AAC speaking community' (Makaton and PECS were, to varying extents, used by everyone in the classroom) involved varying degrees of mastery: staff could be described as AAC 'gatekeepers' who made daily decisions about which laminated symbols would be available, when, and to whom; as well as deciding which Makaton signs would be used and taught within the classroom. Thus, although children used AAC, they were not in the subset of community members who made active decisions about the parameters of AAC usage but rather chose whether or not to deploy what was available (or work around a lack of availability of AAC for their intended meaning by substituting embodied communication strategies, as in the current fragment of data). Saville-Troike (2008) notes, "when a speech event is formalised, there are fewer options for participants; thus, as language becomes more formalised, more social control is exerted on participants" (p.35).

My observations suggest that children encountered significant levels of structure at the snack table, which limited the range of communicative choices available to them. For instance, both the physical environment (the C-shaped table which allowed the leading staff member to face each child in turn) and the functional emphasis on requesting (reflected in the range of PECS symbols provided) both oriented strongly towards a horizontal exchange (staff-student) rather than a vertical exchange (student-student). Since the leading staff member was the gatekeeper to the food and drink and requesting was the encouraged speech function; interaction with peers (or other staff members present) was not foregrounded as relevant to successful enactment of the event.

Luke was a consistently active participant in all recorded observations of snack time: he was very familiar with symbols and could scan them with ease to find his preferred item. He also knew some of the associated Makaton signs and would often attempt to verbalize his request although with variable clarity. In the following transcribed extract, Jane (a Teaching Assistant) is leading snack time. The snack tray has passed to Luke for his third turn at choosing, having previously chosen raisins. Figure 1 depicts the exchange using annotated video stills.

**[INSERT QR Figure 2a]**

[INSERT QR Figure 2b]

**Figure 1. Luke Asks for Raisins: Annotated Video Stills**

### ***Analysis***

In this extract, Luke is firmly rejecting the idea of choosing from the remaining available selection (tomato, apple or carrot); an option which would be easier for him in at least two ways. Firstly, there is the material advantage that symbol cards are available for these items and can be easily deployed in a simple transaction efficient both in terms of time and cognitive effort. Secondly, there is social and transactional benefit associated with providing the expected response which typically involves agreement, acceptance, acquiescence or other validation of the previous speaker's utterance; or as CA literature calls it, a 'preferred response' (Pomerantz, 1984). The established daily routine at snack time in turn derives from the teaching framework associated with PECS implementation (Bondy and Frost, 1994). Whilst the identification of 'preferred' and 'dispreferred' actions is usually established locally in participants' talk, an ethnographic perspective suggests that snack time involves a shared understanding of the expectation that the child will use their allocated turn to lift a symbol card and present it by way of request. Luke therefore performs here a 'dispreferred action': he resists the expectation to select from the available items, and instead chooses to make known his displeasure at the absence of raisins. Performing a dispreferred action has implications for the multimodal orchestration of the act: as the situationally 'legitimated' mode (PECS) permits only acquiescence to the expected routine, resistance requires the use of alternative semiotic resources. Luke achieves this through a complex multimodal orchestration: vocalisations ('Uh?'), verbal imitation ('all gone'), gestural imitation (the upturned palms gesture), gesture (tapping the empty tray space with his finger), direction of gaze (which shifts between Jane's face, Jane's signing hands and the empty tray space), and object manipulation (pulling and lifting the tray). His left hand remaining in resting position in the empty tray space between gestures could be

seen as the gestural equivalent of a 'sound stretch' in verbal conversation: an elongated noise such as *uh* or *em* performed by the speaker to 'hold the floor' whilst they search for the next utterance (Liddicoat, 2011). In this case, the hand remaining in the empty tray space indicates Luke's ongoing orientation towards securing raisins and his wider determination to make himself understood beyond the parameters of available AAC.

To examine how multiple modes are orchestrated together to achieve a communicative goal, Norris (2004) proposes the concepts of *modal intensity* and *modal complexity*. An action which is in the foreground of our attention will possess *modal intensity* (where a single mode can carry the action by itself); or *modal complexity* (many modes are intricately intertwined to produce the action). In this interaction, Luke did not orient towards the usual outcome of requesting through PECS, which carried the risk of Jane concluding that he was disengaging from snack time unless he was able to keep the negotiation open with sufficient modal complexity or intensity. In the following nine second excerpt from the multimodal matrix (Figure 2), an instance of the use of modal complexity emerges:

**[INSERT QR Figure 3]**

### **Figure 2: Luke Asks for Raisins: Extract from Multimodal Matrix**

Here Luke works towards his goal with multiple intertwined modes. His posture orients to the interaction with Jane as he faces her over the desk (and later leans in further); and the questioning function of the rapidly repeated upturned palm gesture combines with the gesturing hand's resting position in the empty raisin space on the tray as a form of deixis, indicating the subject of the questioning. The triadic relationship established between Luke, Jane and the tray (which would normally consist of Luke, Jane and the PECS folder) is established by both the hand gesture and the direction of eye gaze, which alternates regularly between Jane and the tray. Luke vocalises three



times here, in response to Jane's speech: on two occasions with the noise *uh?* and once with a repetition of Jane's utterance, *gone!* Repetition of the interactional partner's prior utterance by an individual with autism is often conceptualised as *echolalia* (Neely et al., 2016), which can pathologise it as a manifestation of disordered speech. However, context-embedded, multimodal analyses of echolalia tend to observe a certain interactive functionality, orderliness and purposefulness in the repetition: for instance, Samuelsson and Ferreira (2013:146) note that the 'recycling' of previous elements of a conversation can constitute 'meaningful contributions to communication'. Here, Luke's repetition of Jane's '*gone!*' is sequentially significant when situated alongside in his multimodal communication at that moment (4:57): direct eye contact with Jane (which is sustained for three seconds, longer than anywhere else in the interaction); ongoing repetition of the upturned palms gesture with a hand that is otherwise resting in the empty tray compartment; and a postural/proxemic orientation to Jane (sitting straight at the desk directly facing her). Luke's 'echolalia' here appears to fulfil multiple functions in the unfolding interaction: it comprises an acknowledgement of the lack of raisins, a demonstration of ongoing orientation to turn-taking and interactional engagement with Jane (performing the expected completion of an 'adjacency pair' through repetition), and the performance of a dispreferred action (declining to perform the expected action of engaging with the symbol cards to choose something else). In this way, Luke succeeds in making his meaning clear by resisting the limited choice made available by the symbol cards and instead orchestrating a range of embodied and idiosyncratic strategies to make an alternative request.

## ***Discussion***

This small fragment of data was examined from three perspectives. The Ethnography of Communication framework contextualised the exchange as a communicative event which was an instantiation of a twice daily communicative situation, with clearly established and mutually understood communicative expectations about who may 'speak'; when; and how. This ethnographic information was significant in determining that Luke's decision to reject the PECS folder and to use embodied communicative strategies constituted a 'dispreferred action' in the wider context of their

activities which extend beyond the transcribed interactions. The EoC framework also permitted critical reflection on the respective positions occupied by Luke and Jane in the 'speech community'; which although bound together by shared understandings of the rules of classroom communication, was also very heterogenous with varying levels of mastery of spoken English and AAC. This is an important contribution to the hybridized approach because it connects to considerations of power and agency, particularly salient issues in the case of disabled research participants (Brewster, 2007). Svennevig et al. (2005) argue that a risk of focusing analytic attention on participants' transcribed talk, such as one might do in CA, is giving the impression of 'a homogenous community, with completely overlapping members' resources' (p.11); where members have near-equal social, cognitive and linguistic power in interaction. Focusing on multimodal microanalysis alone might portray Luke as highly agentic in deploying a range of embodied modes (gaze, vocalisation, object manipulation, touch) to make his request; whilst the EoC framework locates such agentic action within the constraints of community routines, rules and expectations and the finite choice of symbol cards available for communication.

Brewster (2007) points out that AAC can simply serve to replicate existing power relations between the AAC user and staff if only AAC vocabulary deemed institutionally acceptable is provided. Whilst the three symbols made available to Luke do enable him to choose between apple, carrot and tomato, they do not enable him to voice protest, refusal or requests for alternative items or to engage in phatic (social) communicative exchanges. This means that he must by necessity have recourse to non-verbal embodied communication to realize these speech functions. Of course, this is not an inherent or ubiquitous limitation of AAC systems which can comprise comprehensive vocabulary sets. Nevertheless, issues around power, ableism and control in AAC provision (and in interactions between disabled and non-disabled people generally) need to be acknowledged lest the multimodal analysis overstate the agency of the AAC user, when in fact institutional limitations on available vocabulary may constitute powerful constraints on the parameters of the choice in modes to communicate.

As in previous studies involving children with ASD (Dickerson et al., 2007; Stribling et al., 2007; Muskett et al., 2013), concepts from CA have been useful in establishing the functionality and interactional work in Luke's actions, which might otherwise be pathologized as symptoms of autism. For instance, with the appropriation of CA tools it was possible to identify how Luke completed adjacency pairs in a variety of ways including repetition ('echolalia'), vocalisation, and gesture; leaving his hand to rest in the empty space on the snack tray served as a gestural equivalent of a 'sound stretch', performing the interactional work of 'holding the floor'.

While, appropriating CA concepts has been useful in the hybridized approach explored here, one point of divergence has been the format of transcription that does not adopt the Jeffersonian system. Jefferson transcription is well-placed to capture the atypical 'conversations' of minimally verbal participants who distribute the interactional load of their communication primarily or exclusively across gesture, gaze, and object manipulation. If Luke's exchange with Jane had been transcribed thus, very little speech would have been available for transcription, whilst extensive verbal descriptions of embodied actions in parentheses would have been appended to every short utterance. While Luke's actions could have been captured using multimodally oriented CA transcription conventions (e.g., as developed by Mondada), the multimodal matrix provides another alternative. As Norris (2004) contends, if we are theoretically committed to the idea that language should not have *a priori* privileged status as the dominant mode, there is an argument for transcription methods that shift away from logocentrism. The multimodal matrix, which allocates separate and equally sized columns to groups of modes, can provide a basis for the close sequential analysis of interaction with no inherent privileging of any one particular mode. The annotated video stills were used to complement the multimodal matrix, as a 'transvisual' has the effect of foregrounding modes such as posture and proxemics as well as the physical setting and orientation of participants towards each other; with utterances being relegated to the status of annotation. This was an apt approach to represent Luke's multimodal repertoire.

Finally, the hybridized approach drew on elements of Norris' (2004) framework known as Multimodal (Inter)Action Analysis, and its argument that we bring actions to the foreground of our

continuum of attention (and that of our interactional partner) through *modal intensity* and/or *modal complexity*. For instance, Luke had to carefully navigate a course between two possibilities: on the one hand, he did not want to comply with choosing from the available symbol cards which was the expected outcome of the interaction; but on the other hand he did not want to be interpreted as refusing his turn. Maintaining sufficient modal intensity and/or complexity at all points in the interaction Luke sustained the resolution of the request in the foreground for both him and Jane even though the exchange was potentially liable to foreclosure: he maintained the interaction through his postural and gestural orientation, gaze shifting between Jane's face and the tray (and occasionally Jane's hands when she is signing), and the use of both echolalia and vocalisations.

The hybridized approach has provided a multi-perspectival understanding of this small data fragment by combining two forms of microanalysis (one focusing on the sequentiality and orderliness of talk, the other on how modes were deployed in joint modal configurations). This in turn was situated within contextualised understandings of the shared communicative practices of 'snack time' as an established twice-daily communicative situation within a heterogenous speech community. However, drawing upon multiple perspectives on multimodality is not without its difficulties, and the present exploration does not claim to have resolved the tensions and contradictions that might arise. One such tension might be the admissibility of the 'wider-than-sequential context' (Maynard, 2006: 64) in the analysis that moves beyond the transcribed interactions. Despite the challenges, atypical and minimally verbal communicators such as Luke perhaps require us to continue to work across boundaries, and even transgress the parameters of established perspectives, to respond to the complexity involved in rendering visible their interactional competencies.

#### **Acknowledgements:**

I am grateful to Professor Cathy Burnett, Dr Terhi Korkiakangas and Dr Rosie Flewitt for their helpful comments as well as the anonymous reviewers who provided useful feedback on earlier drafts.

#### **Funding Acknowledgement:**

## References

- American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders (5<sup>th</sup> ed.) Washington, DC: American Psychiatric Association.
- Atkinson JM and Heritage J (1984) Preference organization. In Atkinson JM and Heritage J (eds) *Structures of social action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press, pp.53-6.
- Bloch S, Wilkinson R (2004) The understandability of AAC: A conversation analysis study of acquired dysarthria. *Augmentative and Alternative Communication*, 20(4):272-82.
- Bondy AS and Frost LA (1994) The picture exchange communication system. *Focus on Autism and Other Developmental Disabilities*, 9(3): 1-19.
- Brewster SJ (2007) *Asymmetries of power and competence and implications for AAC: interaction between adults with severe learning disabilities and their care staff*. PhD Thesis, University of Birmingham, UK.
- British Educational Research Association (BERA) (2011) Ethical Guidelines for Educational Research. Available at: <https://www.bera.ac.uk/wp-content/uploads/2014/02/BERA-Ethical-Guidelines-2011.pdf> (Accessed 9 January 2017).
- Clarke M and Bloch S (2013) AAC Practices in Everyday Interaction. *Augmentative and Alternative Communication*, 29(1): 1-2.
- Dickerson P, Stribling P and Rae J (2007) Tapping into interaction: How children with autistic spectrum disorders design and place tapping in relation to activities in progress. *Gesture*, 7(3): 271-303.
- Dreyfus SJ (2006) When there is no speech: a case study of the nonverbal multimodal communication of a child with an intellectual disability. PhD Thesis, University of Wollongong, Australia.
- Engelke CR, Higginbotham DJ (2013) Looking to speak: On the temporality of misalignment in interaction involving an augmented communicator using eye-gaze technology. *Journal of Interactional Research in Communication Disorders*, 4(1): 95-122.
- Enninger W (1987) On the organization of sign-processes in an Old Order Amish (OOA) parochial school. *Research on Language and Social Interaction*, 21(1-4): 143-170.
- Erickson F (2010) The neglected listener: issues of theory and practice in transcription. In Streek J (ed) *New Adventures in Language and Interaction*. Amsterdam: Benjamins, pp. 243-256.
- Flewitt R (2006) Using video to investigate preschool classroom interaction: education research assumptions and methodological practices. *Visual Communication*, 5(1): 25-50.
- Flewitt R, Nind M, Payler J (2009) If she's left with books she'll just eat them': Considering inclusive multimodal literacy practices. *Journal of Early Childhood Literacy*, 9(2): 211-33.
- Flewitt R (2011) Bringing ethnography to a multimodal investigation of early literacy in a digital age. *Qualitative Research*, 11(3): 293-310.

- Goodwin C (2011) Contextures of action. In Streek E, Goodwin C and LeBaron C (eds) *Embodied interaction: Language and body in the material world*. Cambridge: Cambridge University Press, pp.182-193.
- Goodwin C (2007) Participation, stance and affect in the organization of activities. *Discourse & Society*, 18(1): 53-73.
- Goodwin C (2000) Practices of seeing visual analysis: An ethnomethodological approach. In Van Leeuwen T and Jewitt C (eds) *The handbook of visual analysis*. London: Sage, pp.157-182.
- Goodwin C and Heritage J (1990) Conversation analysis. *Annual review of anthropology*, 19(1): 283-307.
- Goodwin MH & Goodwin C (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica*, 62(1/2): 51–75.
- Green J and Bloome D (2004) Ethnography and ethnographers of and in education: A situated perspective. In Flood J, Lapp D, Heath SB (eds) *Handbook of research on teaching literacy through the communicative and visual arts*. New York: MacMillan, pp.181-202.
- Gumperz J (1982) *Discourse Strategies*. (Vol.1) Cambridge: Cambridge University Press.
- Heritage J (1984) *Garfinkel and Ethnomethodology*. Cambridge: Polity.
- Hymes D (1972) Toward ethnographies of communication: The analysis of communicative events. *Language and social context*, 21-44.
- Jefferson, G (2004) Glossary of Transcript Symbols with An Introduction. In: Lerner GH (ed) *Conversation analysis: Studies from the First Generation*. PA: John Benjamins Publishing, pp.13-32.
- Jewitt C, Bezemer J and O'Halloran K (2016) *Introducing Multimodality*. London: Routledge.
- Jewitt C (ed) (2009) *The Routledge handbook of multimodal analysis*. London: Routledge.
- Kasari C, Brady N, Lord C and Tager-Flusberg H (2013) Assessing the minimally verbal school-aged child with autism spectrum disorder. *Autism Research*, 6(6): 479-493.
- Korkiakangas, T. (2018/In Press). Communication, gaze, and autism: A multimodal interaction perspective. London: Psychology Press.
- Korkiakangas T and Rae J (2014) The interactional use of eye-gaze in children with autism spectrum disorders. *Interaction Studies*, 15(2):233-259.
- Korkiakangas TK and Rae, JP (2013) Gearing up to a new activity: How teachers use object adjustments to manage the attention of children with autism. *Augmentative and Alternative Communication*, 29(1):83-103.
- Kovarsky D (2016) A Retrospective Look at the Ethnography of Communication Disorders. *Journal of Interactional Research in Communication Disorders*, 7(1): 1-25.
- Kovarsky D, Damico JS, Maxwell M, Panagos J, Prelock P and Keyser H (1988). The ethnography of communication disorders and its contribution to the study of communication disorders. *Seminar presented at the American Speech-Language Hearing Association Convention*. Boston, MA.
- Kress G (2011) Partnerships in research: multimodality and ethnography. *Qualitative Research*, 11(3):239-260.
- Kress G, Van Leeuwen T (2001) *Multimodal Discourse: the Modes and Media of Contemporary Communication*. London: Edward Arnold.
- Lancaster L (2007) Representing the ways of the world: How children under three start to use syntax in graphic signs. *Journal of Early Childhood Literacy*, 7(2): 123-154.

Lerner GH, Zimmerman DH and Kidwell M (2011) Formal structures of practical tasks: A resource for action in the social life of very young children. In Streek E, Goodwin C and LeBaron C (eds) *Embodied interaction: Language and body in the material world*. Cambridge: Cambridge University Press, pp.44-58.

Liddicoat AJ (2011) *An Introduction to Conversation Analysis*. London: Continuum.

Mavers D (2012) Transcribing video. National Centre for Research Methods MODE Working Paper 4, May 2012. Available at: <http://eprints.ncrm.ac.uk/2877/> (Accessed on 1 July 2016).

Maynard DW (2006) Ethnography and Conversation Analysis: What is the Context of an Utterance? In Hesse-Biber SN and Leavy P (2006) *Emergent methods in social research*. London: Sage, pp.55-94.

McHoul A, Rapley M and Antaki C (2008) You gotta light? On the luxury of context for understanding talk in interaction. *Journal of Pragmatics*, 40: 827-839.

Mellman LM, DeThorne LS, Hengst JA (2010) "Shhhh! Alex Has Something To Say": AAC-SGD Use in the Classroom Setting. *Perspectives on Augmentative and Alternative Communication*, 19(4):108-14.

Moerman M (1988) *Talking culture: Ethnography and conversation analysis*. Philadelphia: University of Pennsylvania Press.

Mondada L (2016) Challenges of multimodality: Language and the body in social interaction. *Journal of Sociolinguistics*, 20(3): 336-366.

Mondada L (2011) The organization of concurrent courses of action in surgical demonstrations. In Streek E, Goodwin C and LeBaron C (eds) *Embodied interaction: Language and body in the material world*. Cambridge: Cambridge University Press, pp.207-226.

Muskett T and Body R (2013) The case for multimodal analysis of atypical interaction: Questions, answers and gaze in play involving a child with autism. *Clinical linguistics & phonetics*, 27(10-11): 837-850.

Muskett T, Perkins M, Clegg J and Body R (2010) Inflexibility as an interactional phenomenon: Using conversation analysis to re-examine a symptom of autism. *Clinical linguistics & phonetics*, 24(1): 1-16.

Neely L, Gerow S, Rispoli M, Lang R and Pullen N (2016) Treatment of Echolalia in Individuals with Autism Spectrum Disorder: a Systematic Review. *Review Journal of Autism and Developmental Disorders*, 3(1): 82-91.

Nevile M (2015) The embodied turn in research on language and social interaction. *Research on Language and Social Interaction*, 48(2): 121-151.

Nind M (2008) Conducting qualitative research with people with learning, communication and other disabilities: methodological challenges. Available at: <http://eprints.ncrm.ac.uk/491/> (accessed 1 September 2016).

Norris S (2004) *Analyzing multimodal interaction: A methodological framework*. New York: Routledge.

Ochs E, Kremer-Sadlik T, Sirota K & Solomon O (2004). Autism and the social world: An anthropological perspective. *Discourse Studies* 6(2): 147-183.

O'Halloran K and Smith BA (2012) Multimodal Studies. In O'Halloran K and Smith BA (eds) *Multimodal studies: Exploring issues and domains* (Vol. 2). Abingdon: Routledge.

Pomerantz A (1984) Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes. In Atkinson JM and Heritage J (eds) *Structures of Social Action*. Cambridge: Cambridge University Press, pp.57-101.

- Prizant BM, Wetherby AM, Rubin E and Laurent AC (2003) The SCERTS Model: A transactional, family-centered approach to enhancing communication and socioemotional abilities of children with autism spectrum disorder. *Infants & Young Children*, 16(4): 296-316.
- Rampton B, Roberts C, Leung C and Harris R (2002) Methodology in the analysis of classroom discourse: Response article. *Applied Linguistics*, 23: 373-392.
- Roulstone S, Wren Y, Bakopoulou I, Goodlad S and Lindsay G (2012) Exploring interventions for children and young people with speech, language and communication needs: A study of practice. London: DfE.
- Sacks H, Schegloff EA and Jefferson G (1974) A simplest systematics for the organization of turn-taking for conversation. *Language*, 50: 696-735.
- Samuelsson C, and Ferreira J (2013) Recycling in communication involving a boy with autism using picture exchange system (PECS). In: Norén N, Samuelsson C and Plejert C (eds) *Aided Communication in Everyday Interaction*. Guildford, J&R Press.
- Saville-Troike M (2008) *The ethnography of communication: An introduction*. Hoboken: John Wiley & Sons Ltd..
- Schegloff EA (1992) Repair after next turn: The last structurally provided defense of intersubjectivity in conversation. *American journal of sociology*, 97(5):1295-1345.
- Scollon R (2001) *Mediated discourse: the nexus of practice*. London: Routledge.
- Sheehy K and Duffy H (2009) Attitudes to Makaton in the ages on integration and inclusion. *International Journal of Special Education*, 24(2): 91-102.
- Sigman S (1987) Multichannel communication codes. *Special issue for Research on Language and Social Interaction*, 20.
- Simmons-Mackie N and Damico JS (1999). Social role negotiation in aphasia therapy: Competence, incompetence and conflict. In Kovarsky D, Duchan J and Maxwell M (eds) *Constructing (In)competence: Disabling Evaluations in Clinical and Social Interaction*. Mahwah, NJ: Erlbaum, pp.313-324.
- Solomon O (2008) Language, autism, and childhood: An ethnographic perspective. *Annual Review of Applied Linguistics*, 28: 150-169.
- Stiegler LN (2007) Discovering communicative competencies in a nonspeaking child with autism. *Language, Speech, and Hearing Services in Schools*, 38(4): 400-413.
- Stivers T and Sidnell J (2005). Introduction: multimodal interaction. *Semiotica*, 2005(156): 1-20.
- Stribling P, Rae J and Dickerson P (2007) Two forms of spoken repetition in a girl with autism. *International Journal of Language & Communication Disorders*, 42(4): 427-444.
- Svennevig J and Skovholt K (2005) The methodology of conversation analysis—positivism or social constructivism? In: *9th International Pragmatics Conference*, Riva del Garda, Italy, 10-15 July 2005, pp.10-15.
- Tager-Flusberg H and Kasari C (2013) Minimally verbal school-aged children with autism spectrum disorder: the neglected end of the spectrum. *Autism Research*, 6(6): 468-478.
- Taylor R (2012) Messing about with metaphor: multimodal aspects to children's creative meaning making. *Literacy*, 46(3): 156-166.



Thomas S (1987) Non-lexical soundmaking in audience contexts. *Research on Language and Social Interaction*, 21(1–4): 189–22

Wilkinson R (2013) Conversation analysis and Communication Disorders. In Ball MJ, Perkins MR, Müller N and Howard S (eds) *The Handbook of Clinical Linguistics*. Oxford: Blackwell, pp.92-106

World Health Organisation (1992) *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. Geneva: World Health Organisation.